

## 10.4 ASSESSING THE QUALITY ASSURANCE SYSTEM FOR THE OKLAHOMA MESONET WITH ACCURACY MEASURES

Peter K. Hall Jr., Alexandria G. McCombs, Christopher A. Fiebrich, and Renee A. McPherson\*  
Oklahoma Mesonet, Norman, Oklahoma

### 1. INTRODUCTION

Quality assurance (QA) meteorologists for the Oklahoma Mesonet issue a “trouble ticket” when they detect a problem with a particular sensor (McPherson et al. 2007). The trouble ticket indicates to a Mesonet technician that a repair or replacement is needed for a specific sensor at a specific Mesonet site. Sometimes the resolution of the sensor problem simply requires rewiring or adjustment; other times the sensor needs to be replaced.

Sensors also can be replaced on a pre-assigned basis (i.e., scheduled rotation). The “rotated” sensors are not changed because of a particular problem, but when they have been in the field for a predetermined, sensor-dependent length of time (Fiebrich et al. 2006).

When sensors are replaced, either because of a problem or rotation, the old sensors are returned to the calibration laboratory for an “as found” test. The “as found” test compares the sensor to a reference sensor of similar type. At this point, a lab technician can diagnose if the sensor needs to be reconditioned, repaired, or retired, or if there was no problem.

Clearly, the Mesonet QA staff strives to minimize the number of trouble tickets issued on sensors that do not have problems. The “as found” tests performed on the sensors returning from the field (regardless of why a trouble ticket was issued) are important to the QA staff to determine if a problem was identified correctly or if a rotated sensor had a sensor problem.

Accuracy measures and skill scores typically have been applied to forecast verification (e.g., Mason 1982; Doswell et al. 1990), but were used here to assess the quality assurance process completed by manual methods. This paper focuses on accuracy measures of seven different variables, as calculated from the “as found” sensor tests of the Oklahoma Mesonet.

### 2. ACCURACY MEASURE PROCEDURE

Accuracy measures were calculated on the following variables: air temperature at 1.5 m (TAIR) from a fast-response thermistor (hereafter “fasttherm”), air temperature at 1.5 m (TSLO) from a slower-response thermistor, relative humidity at 1.5 m (RELH), soil temperatures at 5, 10, and 30 cm (SOIL), pressure (PRES), wind speed at 2 or 9 m (WSEN), and wind speed at 10 m (WSPD).

The accuracy measures calculated for these seven variables were based on a dichotomous contingency table (Table 1; Wilks 1995). If a trouble ticket were issued and the sensor failed calibration, a “Hit” occurred. However, if a ticket were issued but the sensor did not fail calibration, a “False Alarm” was counted. A “Miss” constituted a sensor that returned to the lab for rotation but it failed the calibration test (i.e., the QA staff did not detect a confirmed problem). Finally, a “Correct Negative” referred to a sensor that returned for rotation and did not fail the calibration test.

Calibration sheets for calendar year 2007 were analyzed to place the results of a given sensor test in one of the four categories. The sum of each category was used to calculate specific accuracy measures, as detailed in equations 1–5. The total number of events (“Total Number”) also was obtained.

---

\* *Corresponding author address:* Renee A. McPherson, Oklahoma Climatological Survey, University of Oklahoma, 120 David L. Boren Blvd., Suite 2900, Norman, OK 73072– 7305; email: [renee@ou.edu](mailto:renee@ou.edu).

	<b>PROBLEM? YES</b>	<b>PROBLEM? NO</b>
<b>TICKET? YES</b>	Hit	False Alarm
<b>TICKET? NO</b>	Miss	Correct Negative

**Table 1.** Dichotomous contingency table for the evaluation of results of the “as found” sensor tests.

QA meteorologists used five different accuracy measures to analyze the trouble ticket process: Proportion of Correct (PoC), False Alarm Rate, Bias, Probability of Detection (PoD), and Threat Score. PoC indicated the ratio of Hits and Correct Negatives to the Total Number of events (Eq. 1). A perfect score was 1; the worst score was 0.

$$PoC = (Hits + Correct\ Negative) \div Total\ Number \quad (1)$$

False Alarm Rate (FAR) indicated the fraction of tickets issued for sensors without confirmed problems compared to all tickets issued (i.e., ratio of incorrect ticketing; Eq. 2). Opposite of PoC, 0 was a perfect score for FAR while 1 was the worst score.

$$FAR = False\ Alarms \div (Hits + False\ Alarms) \quad (2)$$

To examine if there was over- or under-ticketing of sensors, Bias was calculated (Eq. 3). If the score equaled 1, there was no bias. Scores less than 1 denoted under-ticketing; those greater than 1 revealed over-ticketing.

$$Bias = (Hits + False\ Alarms) \div (Hits + Misses) \quad (3)$$

Probability of Detection (PoD) indicated the fraction of tickets that were issued when there was a sensor problem (Eq. 4). A perfect PoD was 1; the worst score was 0.

$$PoD = Hits \div (Hits + Misses) \quad (4)$$

The Threat Score identified tickets issued when there was an actual problem (Eq. 5). Unlike PoC, the Threat Score did not account for Correct Negatives. A perfect score was 1; the worst score was 0.

$$Threat\ Score = Hits \div (Hits + False\ Alarms + Misses) \quad (5)$$

### 3. RESULTS

Table 2 summarizes the number of “as found” sensor tests (n), hits (YY), false alarms (YN), misses (NY), and correct negatives (NN) by variable for 2007. The numbers were used to calculate the accuracy measures listed in section 2.

Variable	n	YY	YN	NY	NN
TAIR	23	6	3	2	12
TSLO	102	3	4	4	91
RELH	101	12	1	2	86
SOIL	101	73	14	7	7
PRES	53	0	1	1	51
WSEN	108	8	4	33	63
WSPD	18	2	2	1	13

**Table 2.** List of the seven variables examined: air temperature at 1.5 m (TAIR) from a fast-response thermistor, air temperature at 1.5 m (TSLO) from a slower-response thermistor, relative humidity at 1.5 m (RELH), soil temperatures at 5, 10, and 30 cm (SOIL), pressure (PRES), wind speed at 2 or 9 m (WSEN), and wind speed at 10 m (WSPD). The columns denote, by variable, the total number of “as found” sensor tests (n), hits (YY), false alarms (YN), misses (NY), and correct negatives (NN).

#### 3.1. Proportion of Correct

PoCs for all variables (Fig. 1) except 2- or 9-m wind speed (WSEN) were at least 0.75. The lowest PoC was 0.66 (for the wind sentry) and the highest was 0.97 (for relative humidity). Most of the values were dominated by Correct Negatives, potentially inflating the PoC. Only the soil temperature sensors (SOIL) had a small number of Correct Negatives (7% of all tickets) because, unlike all other sensors, there was no standard procedure for soil sensor rotation. Thus, more than 70% of the tested sensors that counted as hits resulted from sensor rotations.

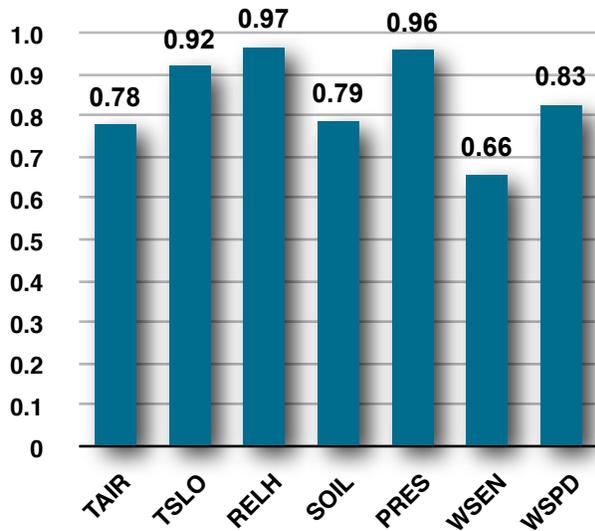


Fig. 1. Probability of Correct values for select Mesonet variables. The seven variables examined were as follows: air temperature at 1.5 m (TAIR) from a fast-response thermistor, air temperature at 1.5 m (TSLO) from a slower-response thermistor, relative humidity at 1.5 m (RELH), soil temperatures at 5, 10, and 30 cm (SOIL), pressure (PRES), wind speed at 2 or 9 m (WSEN), and wind speed at 10 m (WSPD).

### 3.2. False Alarm Ratio

After PoC, the scores were not as uniform for the other tests. The FAR (Fig. 2) varied from 0.08 to 1.0. The relative humidity sensors had only one false alarm, thus its FAR was 0.08%. The soil temperature sensors also had a low FAR (0.16). Pressure had the highest false alarm rate, as there was only one ticket issued during the year, and the barometer tested fine during the “as found” calibration.

### 3.3. Bias

The Bias measure can indicate where more attention to sensor problems is warranted. Figure 3 displays the Bias of each variable. Generally, there was no bias for the slower-response air temperature, soil temperature, or pressure measurements. Relative humidity and wind sentries at 2 and 9 m were under-ticketed; however, the fasttherms and wind monitor nose cones

were over-ticketed. To try to reduce bias, more attention will be paid to the fasttherms (13% over-ticketing) and wind sentries (71% under-ticketing) in the future.

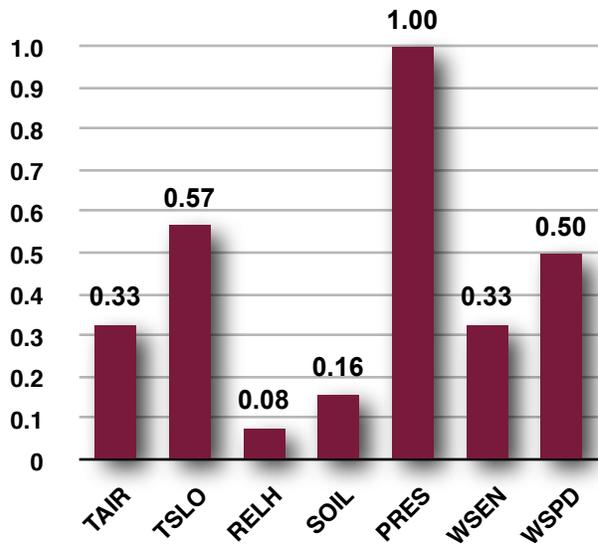


Fig. 2. As in Fig. 1 except for False Alarm Rate.

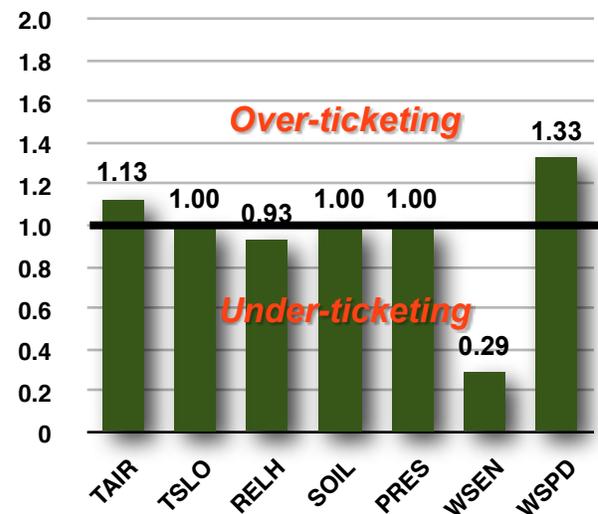


Fig. 3. As in Fig. 1 except for Bias.

### 3.4. Probability of Detection

The best PoD scores ( $\geq 0.75$ ) were for soil temperature, relative humidity, and air temperature measured by the fasttherms (Fig. 4). Pressure sensors and wind sentries had the worst PoD values at 0.0 and 0.20,

respectively. The PoD score was influenced by the number of Misses (assuming there were some Hits). If a sensor returned from the field because of a scheduled rotation, but failed calibration, it was counted as a miss. Thus, a variable with a low PoD highlighted that the sensor needed to be examined more closely near the end of its rotation period or perhaps even that the rotation schedule needed to be adjusted. In the case of the wind sentries (Fig. 4), the rotation time may need to be decreased.

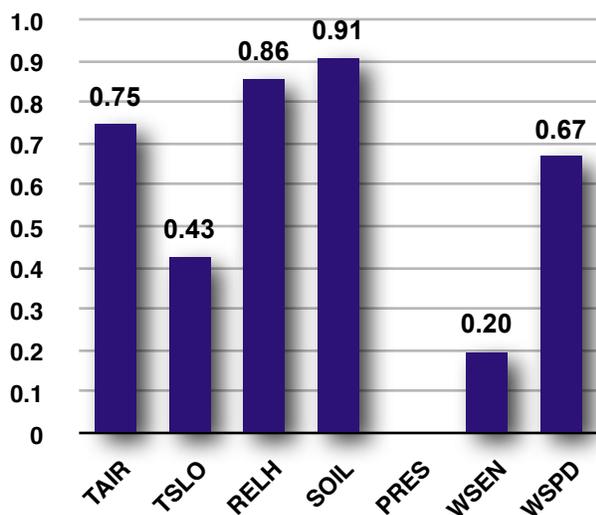


Fig. 4. As in Fig. 1 except for Probability of Detection.

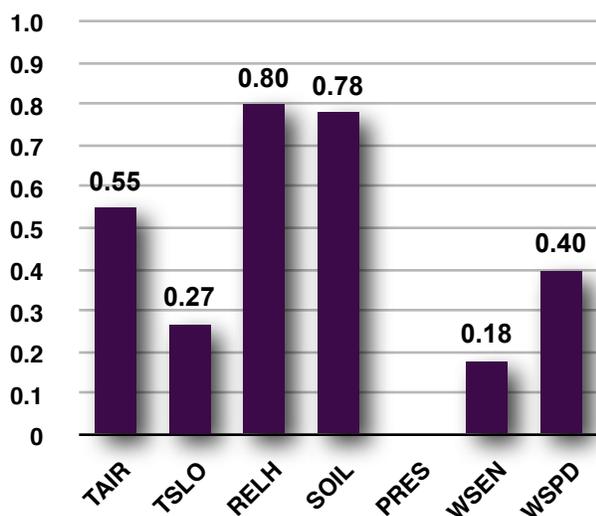


Fig. 5. As in Fig. 1 except for Threat Score.

### 3.5. Threat Score

The Threat score allowed for the exclusion of non-events (Correct Negatives). Thus, Figure 5 illustrates the likelihood of finding a problem when an event occurred. The highest scores were those associated with soil temperature (0.78) and relative humidity (0.83) measurements, whereas the worst score was associated with pressure sensors (0.0) and wind sentries (0.18). All aboveground thermistors and wind monitor nose cones had values ranging from 0.27 to 0.55.

## 4. SUMMARY

The calculation of accuracy measures enhanced the evaluation of the Oklahoma Mesonet's manual quality assurance procedures. The QA meteorologists had their greatest successes identifying problems with relative humidity sensors. The sensor that offered the greatest challenge, in terms of PoC, was the wind sentry (wind speed at 2 and 9 m).

FAR values demonstrated that tickets on fasttherms, slower-response thermistors, barometers, and wind monitors need to be examined more carefully. It was found that the three tickets issued on fasttherms were not problems after all. It is assumed that true meteorological conditions caused the  $\sim 0.5^{\circ}\text{C}$  anomalies that caused the QA meteorologist to issue the trouble ticket. Barometers provide the greatest challenge to decreasing false alarms because pressure problems are rare and identifying a  $>0.4$  hPa error in field data is challenging. More development of tests and a greater understanding of the site microclimates could help decrease the FAR for wind speeds at 2, 9, and 10 m.

The wind sentries were significantly under-ticketed by QA meteorologists during this study period. A greater understanding of the sensor should result in increased probability of detection in the future. More stringent analysis will be needed to detect more problems while the sensor is in the field.

The PoD of sensor problems varied greatly from sensor to sensor. The PoD could be considered the QA staff's confidence of

the ticket issued truly being a problem. Thus, there is strong confidence in tickets issued on soil temperature, relative humidity, fasttherms, and wind speed sensors. The remaining sensors have 50% or less confidence of actually being a problem.

Finally, the Threat score results were good for relative humidity and soil temperature, but poor for the remaining sensors. Increasing the hits while decreasing the misses will allow the other sensors to have a greater Threat Score.

Soil temperature measurements, in particular, were associated with relatively high PoC (0.79), PoD (0.91), and Threat Score (0.78) as well as low FAR (0.16) and no Bias (1.0). These values appeared good because the field technician completed an in-field calibration check on each sensor when a ticket was issued. These in-field tests greatly reduced the number of sensors that were returned to the calibration laboratory when there was no problem (False Alarms). Because this procedure seemed successful for the soil temperature sensors, Oklahoma Mesonet QA meteorologists and field technicians seek to test other sensors in the field in a similar manner.

Although reducing residence time for each sensor type could reduce the number of misses, a shorter rotation schedule increases maintenance costs. Hence, the Mesonet QA staff will seek to change residence times only for those sensors with low PoDs, such as wind sentries (Fig. 4).

Now that accuracy measures have been examined for 2007, the Mesonet QA meteorologists plan to check these values on an annual basis. Monitoring these performance metrics should help develop new problem detection techniques for each calibrated sensor, leading to better detection of real sensor problems and fewer incorrectly identified problems.

## 5. ACKNOWLEDGEMENTS

Oklahoma's taxpayers fund the Oklahoma Mesonet through the Oklahoma State Regents for Higher Education and Oklahoma Department of Public Safety. The authors thank the field, laboratory, and quality assurance staff of the Oklahoma Mesonet for their dedication to obtaining the highest quality data possible for the data users.

## 6. REFERENCES

- Doswell, C. A., III, R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585.
- Fiebrich, C. A., D. L. Grimsley, R. A. McPherson, K. A. Kesler, and G. R. Essenberg, 2006: The value of routine site visits in managing and maintaining quality data from the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **23**, 406-416.
- Mason I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303
- McPherson, R. A., C. Fiebrich, K. C. Crawford, R. L. Elliott, J. R. Kilby, D. L. Grimsley, J. E. Martinez, J. B. Basara, B. G. Illston, D. A. Morris, K. A. Kloesel, S. J. Stadler, A. D. Melvin, A.J. Sutherland, and H. Shrivastava, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *J. Atmos. Oceanic Tech.*, **24**, 301-321.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.